

Weekly Report

Lu Junhua

2016 年 12 月 19 日

Done

- Learned about what has been done in Yuquan. They didn't do much on data cleansing, e.g., removal of redundancy, semantically conflicts, etc. Their current strategy is simple. Besides, we discussed about potential ideas about detection.
- Brainstorm on several ideas of vis.
- Collected and read materials on deep learning and GPU acceleration, prepared for the course essay (Advanced Computer Architecture).

To do

- Start writing course essay(Advanced Computer Architecture).
- Prepare questions to ask in Netease, and take internship there.

一些想法

- 事件(时间序列)可视化→单纯找异常特征
对各种相关行为序列进行可视化
对不同组、不同类别玩家的序列进行对比
预先做一些聚类, 分类展现序列
- 序列相似性展示,提供上下文解释相似性→发现可能异常并分析是否真的是异常以及异常原因
相似性是一个或者多个的指标, 但是具体为什么相似或者不相似, 还是需要上下文来解释
多种技术结合(不同视图, 互相连动)
- 结合异常检测模型交互可视分析
寻找特征而不是预先自定义特征(TargetVue相当于是预定义特征)
考虑到外挂一般都是工作室操纵的,往往同一个挂表现出来的总有高度相似性, 并且可能有更频繁的操作. 利用一种张量模型, 检测张量中的密集块.
模型的输入是不同特征, 以及不同特征初始取值区间, 作为种子
不断进行迭代找出最终结果
上两步以后得到一些结果, 一方面又可以用一个视图作为结合上下文进行分析; 一方面与ground truth比较, 还可以交互改进模型(初始种子的改进).

- 异常的异常的交互可视分析

假设设定某种发掘规则, 去发掘异常

对发掘出异常进行量化, 比如找到了多少异常?

对量化值进行观测, 是否有突变? 有的话说明要改变规则, 然后回到上面循环.

- 以上各点有些其实可以彼此结合起来. 此外他们都是基于特征或者规则可变的假设(就是说开发外挂的人, 外挂也一直在针对性的升级). 文章需要考虑的一个大问题是贡献点在哪, 如果说只是数据新颖, 以及没有用可视化解决这个问题, 不知道够不够.

论文

- CoRR 2016 *Visual Analysis of Hidden State Dynamics in Recurrent Neural Networks* 针对RNN提出的一个神经网络可视化系统LSTMVis. RNN对时间序列(文本似乎也被当成了时间序列)的隐藏特征有奇效, 文章举的例子也是文本. 一定程度上来说采用的编码是非常简单的, 没有酷炫的设计, 但是从头到尾确实是在做隐藏特征分析与解读这一件事情. 分析解读以后, 还能找到类似特征在文本中的出现情况并作展示. 这样的系统设计需要比较扎实的基础知识.
- VAST 2014 *VAET: A Visual Analytics Approach for E-transactions Time-Series* 谢聪师兄的老文章. 遵循从总到分, 交互改进的套路. KnotLine的可视设计比较美观, 类似于小glyph设计, 但是真的数据量很大的时候也说不定效果会怎样(要么用聚合, 不然每个视图都承受不了大数据量的感觉). 编码具有一定的灵活性, 这是比较有趣的. 此外, 谢聪师兄的文章似乎总能把小问题描述的相对完备, 这一点值得学习.
- Journal of Biomedical Informatics 2014 *A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data* David Gotz那边的Event可视化文章. 文中对序列进行了简单的类似关联规则那样的挖掘, 挖掘的交互成分主要在指定最小的序列长度值以及support部分. 其散点图的三步动画transition, 用于比较时变的trend是一种很好的方法, 在这个分析的情境下, 对于用户来说是很实用的.
- 一些相关于经济行为的小文章, 提出了两种外挂掘金卖金, 以及有人收金在卖的网络结构. 这个问题下周我也会去网易咨询, 因为之前说过没有聊天记录, 但是这种买卖有时候也能看作一种社交.
- KDD 2016 *Modeling Precursors for Event Forecasting via Nested Multi-Instance Learning* 预测事件将要发生, 以及造成事件的可能原因. 将事件分成不同的层次结构, 在不同层上做约束预测, 最小化损失函数. 并用在线随机梯度下降进行解. 文章比较好的地方可能是能够对单个事件的威胁指数进行计算, 而不是只对某一天或者某一片的事件计算. 相当于不同粒度的可视化, 以后也许会有地方可以用到.